OIPE JAN 1 9 2007 PATENT & TRADEMARK OFFICE

PTO/SB/21 (09-04)
Approved for use through 07/31/2006. OMB 0651-0031
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Application Number	10/601,084
Filing Date	June 20, 2003
First Named Inventor	Hooykaas et al.
Art Unit	1636
Examiner Name	J. Duntson
Attorney Docket Number	2183-6028US

Total Number of Pages in This Submission: 25

ENCLOSURES *(check all that apply)*

- ☐ Fee Transmittal Form
  - ☐ Fee Attached
- ☐ Amendment / Reply
  - ☐ After Final
  - ☐ Affidavits/declaration(s)
- ☐ Extension of Time Request
- ☐ Express Abandonment Request
- ☐ Information Disclosure Statement
- ☒ Certified Copy of Priority Document(s) (EP00204693.6)
- ☐ Reply to Missing Parts/ Incomplete Application
  - ☐ Reply to Missing Parts under 37 CFR1.52 or 1.53
- ☐ Drawing(s)
- ☐ Licensing-related Papers
- ☐ Petition
- ☐ Petition to Convert to a Provisional Application
- ☐ Power of Attorney, Revocation Change of Correspondence Address
- ☐ Terminal Disclaimer
- ☐ Request for Refund
- ☐ CD, Number of CD(s) ____
  - ☐ Landscape Table on CD
- ☐ After Allowance Communication to TC
- ☐ Appeal Communication to Board of Appeals and Interferences
- ☐ Appeal Communication to TC (**Appeal Notice, Brief, Reply Brief**)
- ☐ Proprietary Information
- ☐ Status Letter
- ☐ Other Enclosure(s) *(please identify below)*:

Remarks

The Commissioner is authorized to charge any additional fees required but not submitted with any document or request requiring fee payment under 37 C.F.R. §§ 1.16 AND 1.17 TO Deposit Account 20-1469 during pendency of this application.

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm	TraskBritt, P.C.
Signature	Allen C Turner
Printed Name	Allen C. Turner
Date	January 8, 2007
Reg. No.	33,041

CERTIFICATE OF TRANSMISSION/MAILING

I hereby certify that this correspondence is being facsimile transmitted to the USPTO or deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date shown below.

Signature	ABlackburn
Typed or printed name	Aubry Blackburn
Date	January 8, 2007

Europäisches Patentamt | European Patent Office | Office européen des brevets

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. | Patent application No. | Demande de brevet n°

00204693.6

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

I.L.C. HATTEN-HECKMAN

Anmeldung Nr:
Application no.: 00204693.6
Demande no:

Anmeldetag:
Date of filing: 22.12.00
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Universiteit Leiden
Stationsweg 46
2312 AV Leiden
PAYS-BAS
Stichting Binair Vector Systeem
Floris versterlaan 12
2343 RS Oegstgeest
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Nucleic acid integration in eukaryotes

In Anspruch genommene Prioriät(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

/00.00.00/

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12N15/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR

EPO - DG 1

22. 12. 2000

(60)

Title: Nucleic acid integration in eukaryotes

The invention relates to the field of molecular biology and cell biology. It particularly relates to methods to direct integration towards homologous recombination and uses thereof. Several methods are known to transfer nucleic acids to, in particular, eukaryotic cells. In some methods the nucleic acid of interest is transferred to the cytoplasm of the cell, in some the nucleic acid of interest is integrated into the genome of the host. Many different vehicles for transfer of the nucleic acid are known. For different kinds of cells, different systems can be used, although many systems are more widely applicable than just a certain kind of cells. In plants, e.g., a system based on *Agrobacterium tumefaciens* is often applied. This system is one of the systems that can be used according to the invention.

The soil bacterium *Agrobacterium tumefaciens* is able to transfer part of its tumor-inducing (Ti) plasmid, the transferred (T-) DNA, to plant cells. This results in crown gall tumor formation on plants due to expression of *onc*-genes, which are present on the T-DNA. Virulence (*vir*) genes, located elsewhere on the Ti-plasmid, mediate T-DNA transfer to the plant cell. Some Vir proteins accompany the T-DNA during its transfer to the plant cell to protect the T-DNA and to mediate its transfer to the plant nucleus. Once in the plant nucleus, the T-DNA is integrated at a random position into the plant genome (reviewed by Hooykaas and Beijersbergen 1994, Hansen and Chilton 1999). Removal of the *onc*-genes from the T-DNA does not inactivate T-DNA transfer. T-DNA, disarmed in this way, is now the preferred vector for the genetic modification of plants.

Although much is known about the transformation process, not much is known about the process by which the T-DNA is integrated into the plant genome. It is likely that plant enzymes mediate this step of the transformation process (Bundock et al. 1995). The integration pattern of T-DNA in transformed plants has been extensively studied (Matsumoto et al. 1990, Gheysen et al. 1991, Mayerhofer et al. 1991). The results indicated that T-

DNA integrates via illegitimate recombination (IR) (also called non-homologous recombination, both terms may be used interchangeable herein), a process which can join two DNA molecules that share little or no homology (here the T-DNA and plant target DNA). Even T-DNA molecules in which a large segment of homologous plant DNA was present, integrated mainly by IR and only with very low frequency (1:10^4–10^5) by homologous recombination (HR) (Offringa et al. 1990).

Recently, it was shown that *Agrobacterium* is not only able to transfer its T-DNA to plant cells, but also to other eukaryotes, including the yeast *S.cerevisiae* (Bundock et al. 1995) and a wide variety of filamentous fungi (deGroot et al. 1998). In *S.cerevisiae*, T-DNA carrying homology with the yeast genome integrates via HR (Bundock et al. 1995). However, T-DNA lacking any homology with the *S.cerevisiae* genome becomes integrated at random positions in the genome by the same IR process as is used in plants (Bundock and Hooykaas 1996). Apparently, eukaryotic cells have at least two separate pathways (one via homologous and one via non-homologous recombination) through which nucleic acids (in particular of course DNA), can be integrated into the host genome. The site of integration into a host cell genome is important with respect to the likelihood of transcription and/or expression of the integrated nucleic acid. The present invention provides methods and means to direct nucleic acid integration to a predetermined site through steering integration towards the homologous recombination pathway. The present invention arrives at such steering either by enhancing the HR pathway, or by inhibiting (meaning reducing) the IR pathway.

Host factors involved in the integration of nucleic acid by IR have so far not been identified. The present invention discloses such factors which enables the design of methods for their (temporary) inhibition, so that integration of nucleic acid by IR is prevented, shifting the integration process towards HR and facilitating the isolation of a host cell with nucleic acid integrated by HR at a predetermined site. This is extremely important, since there is no method

available yet for easy and precise genetic modification of a host cell using HR (gene targeting). Of course the actual site of integration is then determined by homology of the nucleic acid of interest with said site.

In a first embodiment the invention provides a method to direct nucleic acid integration of a nucleic acid of interest to a pre-determined site, whereby said nucleic acid has homology at or around the said pre-determined site, in an eukaryote with a preference for non-homologous recombination comprising steering an integration pathway towards homologous recombination. Integration is a complex process wherein a nucleic acid sequence becomes part of the genetic material of a host cell. One step in the process of nucleic acid integration is recombination; via recombination nucleic acid sequences are exchanged and the introduced nucleic acid becomes part of the genetic material of a host cell. In principle two different ways of recombination are possible: homologous and illegitimate or non-homologous recombination. Most (higher) eukaryotes do not or at least not significantly practise homologous recombination although the essential proteins to accomplish such a process are available. One reason for this phenomenon is that frequent use of homologous recombination in (higher) eukaryotes could lead to undesirable chromosomal rearrangements due to the presence of repetitive nucleic acid sequences. To accomplish homologous recombination via a method according to the invention, it is important to provide a nucleic acid which has homology with a pre-determined site. It is clear to a person skilled in the art that the percentage of homology and the length of (a) homologous region(s) play an important role in the process of homologous recombination. The percentage of homology is preferably close to 100%. A person skilled in the art is aware of the fact that lower percentage of homology are also used in the field of homologous recombination, but dependent on, for example, the regions of homology and their overall distribution, lead to a lower efficiency of homologous recombination but are still useful and therefore included in the present invention. Furthermore, the length of a (nearly) homologous region is

approximately 3 kb which is sufficient to direct homologous recombination. At least one homologous region is necessary for recombination but more preferably 2 homologous regions flanking the nucleic acid of interest are used for targeted integration. The researcher skilled in the art knows how to select the proper percentage of homology, the length of homology and the amount of homologous regions. By providing such a homology a nucleic acid is integrated at every desired position within the genetic material of a host cell. It is clear to a person skilled in the art that the invention as disclosed herein is used to direct any nucleic acid (preferably DNA) to any pre-determined site as long as the length of homology and percentage of homology are high enough to provide homologous recombination. A pre-determined site is herein defined as a site within the genetic material contained by a host cell to which a nucleic acid with homology to this same site is integrated with a method according to the invention. It was not until the present invention that a nucleic acid is integrated at every desired position and therefore a method according to the invention is applied, for example, to affect the gene function in various ways, not only for complete inactivation but also to mediate changes in the expression level or in the regulation of expression, changes in protein activity or the subcellular targeting of an encoded protein. Complete inactivation, which can usually not be accomplished by existing methods such as antisense technology or RNAi technology (Zrenner et al, 1993), is useful for instance for the inactivation of genes controlling undesired side branches of metabolic pathways, for instance to increase the quality of bulk products such as starch, or to increase the production of specific secondary metabolites or to inhibit formation of unwanted metabolites. Also to inactivate genes controlling senescence in fruits and flowers or that determine flower pigments. Replacement of existing regulatory sequences by alternative regulatory sequences is used to alter expression of in situ modified genes to meet requirements, (e.g. expression in response to particular physical conditions such as light, drought or pathogen infection, or in response to chemical

inducers), or depending on the developmental state (e.g. in a storage organ, or in fruits or seeds) or on tissue or cell types. Also a method according to the invention is used to effectuate predictable expression of transgenes encoding novel products, for example by replacing existing coding sequences of genes giving a desired expression profile by those for a desired novel product. For example to produce proteins of medicinal or industrial value in the seeds of plants the coding sequence of a strongly expressed storage protein may be replaced by that of the desired protein. As another example existing coding sequences are modified so that the protein encoded has optimized characteristics for instance to make a plant herbicide tolerant, to produce storage proteins with enhanced nutritional value, or to target a protein of interest to an organelle or to secrete it to the extracellular space. As yet another example eukaryotic cells (including yeast, fungi, plant and mammalian cells) are provided with a gene encoding a protein of interest integrated into the genome at a site which ensures high expression levels. As another example the nucleic acid of interest can be part of a gene delivery vehicle to deliver a gene of interest to a eukaryotic cell *in vitro* or *in vivo*. In this way a defect p53 can be replaced by an intact p53. In this way a tumoricidal gene can be delivered to a pre-determined site present only in e.g. proliferating cells, or present only in tumor cells, e.g. to the site where a tumor antigen is expressed from. Gene delivery vehicles are well known in the art and include adenoviral vehicles, retroviral vehicles, non-viral vehicles such as liposomes, etc. As another example the invention is used to produce transgenic organisms. Knock-out transgenics are already produced by homologous recombination methods. The present invention improves the efficiency of such methods. Also transgenics with desired properties are made.

In another embodiment the invention provides a method to direct nucleic acid integration to a pre-determined site, whereby said nucleic acid has homology at or around the said pre-determined site, in an eukaryote with a preference for non-homologous recombination comprising steering an

integration pathway towards homologous recombination by providing a mutant of a component involved in non-homologous recombination. Methods to identify components involved in non-homologous recombination are outlined in the present description wherein *S.cerevisiae* was used as a model system. To this end several yeast derivatives defective for genes known to be involved in various recombination processes were constructed and the effect of the mutations on T-DNA integration by either HR or IR was tested. The results as disclosed herein show that the proteins encoded by *YKU70, RAD50, MRE11, XRS2, LIG4* and *SIR4* play an essential role in DNA integration by IR but not by HR. It is clear to a person skilled in the art that different mutants of a component involved in non-homologous recombination exist. Examples are deletion mutants, knock-out (for example via insertion) mutants or point mutants. Irrespective of the kind of mutant it is important that a component involved in non-homologous recombination is no longer capable or at least significantly less capable to perform it's function in the process of non-homologous recombination. As disclosed herein disruption of *YKU70, RAD50, MRE11, XRS2, LIG4* and *SIR4* did not affect the frequency of DNA integration by HR, showing that these genes are not involved in DNA integration by HR, but only in DNA integration by IR. In another embodiment the invention provides a method to direct integration of a nucleic acid of interest to a sub-telomeric and/or telomeric region in an eukaryote with a preference for non-homologous recombination by providing a mutant of a component involved in non-homologous recombination. A telomeric region is defined herein as a region containing repetitive sequences which is located at the end of a chromosome. Sub-telomeric region is herein defined as a region flanking the telomeric region. As an example is disclosed herein that in yeast strains carrying disruptions of *RAD50, MRE11* or *XRS2* the distribution of integrated DNA copies is altered when compared to wildtype. DNA becomes preferentially integrated in telomeres or subtelomeric regions in the *rad50, mre11* and *xrs2* mutants. A great advantage of integration of DNA copies in telomeres or

subtelomeric regions instead of integration elsewhere in the genomic material is that there is no danger for host genes being mutated or inactivated by a DNA insertion. When in plants deficient for *RAD50*, *MRE11* or *XRS2* DNA copies also integrate into telomeres or telomeric regions, such plants are used for telomeric targeting of T-DNA in transformation experiments to prevent additional insertion mutations from random T-DNA integration into the plant genome.

In yet another embodiment the invention provides a method to direct nucleic acid integration to a pre-determined site, whereby said nucleic acid has homology at or around the said pre-determined site, in an eukaryote with a preference for non-homologous recombination comprising steering an integration pathway towards homologous recombination by partially or more preferably completely inhibiting a component involved in non-homologous recombination. Partial or complete inhibition of a component involved in non-homologous recombination is obtained by different methods, for example by an antibody directed against such a component or a chemical inhibitor or a protein inhibitor or peptide inhibitor or an antisense molecule or an RNAi molecule. Irrespective of the kind of (partial or more preferably complete) inhibition it is important that a component involved in non-homologous recombination is no longer capable or at least significantly less capable to perform it's function in the process of non-homologous recombination. In yet another embodiment the invention provides a method to direct integration of a nucleic acid of interest to a sub-telomeric and/or telomeric region in an eukaryote with a preference for non-homologous recombination by partially or more preferably completely inhibiting a component involved in non-homologous recombination.

In a preferred embodiment the invention provides a method to direct nucleic acid integration to a pre-determined site or to a sub-telomeric and/or telomeric region by providing a mutant of a component involved in non-homologous recombination or by partially or more preferably completely

inhibiting a component involved in non-homologous recombination wherein said component comprises *ku70, rad50, mre11, xrs2, lig4* or *sir4*. Components involved in non-homologous recombination are identified as outlined in the present description. The nomenclature for genes as used above is specific for yeast. Because the nomenclature of genes differs between organisms a functional equivalent or a functional homologue (see for example figure 2 to 5) and/or a functional fragment thereof, all defined herein as being capable of performing (in function, not in amount) at least one function of the yeast genes *ku70, rad50, mre11, xrs2, lig4* or *sir4* are also included in the present invention. A mutant of a component directly associating with a component involved in non-homologous recombination or (partial or complete) inhibition of a component directly associating with a component involved in non-homologous recombination is also part of this invention. Such a component directly associating with a component involved in non-homologous recombination is, for example, identified in a yeast two hybrid screening. An example of a component directly associating with a component involved in non-homologous recombination is *KU80*, which forms a complex with *KU70*. In a more preferred embodiment the invention provides a method to direct nucleic acid integration in yeast, fungus, plant or animal.

In another embodiment the invention provides a method to direct nucleic acid integration to a pre-determined site, whereby said nucleic acid has homology at or around the said pre-determined site, in an eukaryote with a preference for non-homologous recombination comprising steering an integration pathway towards homologous recombination by transiently (partially or more preferably completely) inhibiting integration via non-homologous recombination. In yet another embodiment the invention provides a method to direct integration of a nucleic acid of interest to a sub-telomeric and/or telomeric region in an eukaryote with a preference for non-homologous recombination by transiently (partially or more preferably completely) inhibiting integration via non-homologous recombination. In a more preferred

embodiment such a method is used for yeast, plant or fungus and the transient (partial or more preferable complete) inhibition is provided by an *Agrobacterium* Vir-fusion protein capable of (partially or more preferably completely) inhibiting a component involved in non-homologous recombination or capable of (partially or more preferably completely) inhibiting a functional equivalent or homologue thereof or capable of (partially or more preferably completely) inhibiting a component directly associating with a component involved in non-homologous recombination. In a even more preferred embodiment such a *Agrobacterium* Vir fusion protein comprises VirF or VirE2. It was shown that the *Agrobacterium* VirF and VirE2 proteins are directly transferred from *Agrobacterium* to plant cells during plant transformation (Vergunst et al. 2000). To, for example, accomplish T-DNA integration by HR in plants, VirF fusion proteins containing for example a peptide inhibitor of IR in plant cells are introduced concomitantly with the targeting T-DNA. It has been reported that the C-terminal part (approximately 40 amino acids) of VirF or VirE2 is sufficient to accomplish transfer of T-DNA. A functional fragment and/or a functional equivalent of VirF or VirE is therefore also included in the present invention. In an even more preferred embodiment a component involved in non-homologous recombination comprises *ku70, rad50, mre11, xrs2, lig4* or *sir4* or functional equivalents or homologous thereof or associating components. The nomenclature for genes as used above is specific for yeast. Because the nomenclature of genes differs between organisms a functional equivalent or a functional homologue (see for example figure 2 to 5) and/or a functional fragment thereof, all defined herein as being capable of performing (in function, not in amount) at least one function of the yeast genes *ku70, rad50, mre11, xrs2, lig4* or *sir4* are also included in the present invention. By transiently (partially or more preferably completely) inhibiting a component involved in non-homologous recombination a nucleic acid is integrated at any position without permanently modifying a component involved in non-homologous recombination and preventing unwanted side

effects caused by the permanent presence of such a modified component involved in non-homologous recombination.

Methods according to the present invention, as extensively discussed above, are used in a wide variety of applications. One embodiment of the present invention is the replacement of an active gene by an inactive gene according to a method of the invention. Complete inactivation, which can usually not be accomplished by existing methods such as antisense technology or RNAi technology, is useful for instance for the inactivation of genes controlling undesired side branches of metabolic pathways, for instance to increase the quality of bulk products such as starch, or to increase the production of specific secondary metabolites or to inhibit formation of unwanted metabolites. Also to inactivate genes controlling senescence in fruits and flowers or that determine flower pigments. Another embodiment of the present invention is the replacement of an inactive gene by an active gene. One example is the replacement of an defect p53 by an intact p53. Many tumors acquire a mutation in p53 during their development which renders it inactive and often correlates with a poor response to cancer therapy. By replacing the defect p53 by an intact p53, for example via gene therapy, conventional anti cancer therapy have better changes of succeeding. In yet another embodiment of the invention a therapeutic proteinaceous substance is integrated via a method of the invention. In this way a tumoricidal gene can be delivered to a pre-determined site present only in e.g. proliferating cells, or present only in tumor cells, e.g. to the site where a tumor antigen is expressed from. In yet another embodiment the invention provides a method to introduce a substance conferring resistance for an antibiotic substance to a cell according to a method of the invention. Also a method according to the invention is used to confer a desired property to an eukaryotic cell. In an preferred embodiment method a gene delivery vehicle is used to deliver a desired nucleic acid to a pre-determined site. Gene delivery vehicles are well known in the art and include adenoviral vehicles, retroviral vehicles, non-viral vehicles such as

liposomes, etc.. In this way, a for example, tumoricidal gene can be delivered to a pre-determined site present only in e.g. proliferating cells, or present only in tumor cells, e.g. to the site where a tumor antigen is expressed from.

Furthermore a method according to the invention is used to improve gene targeting efficiency. Such a method is used to improve for example the gene targeting efficiency in plants. In plants transgenes integrate randomly into the genome by IR (Mayerhofer et al. 1991, Gheysen et al. 1991). The efficiency of integration by HR is very low, even when large stretches of homology between the transgene and the genomic target site are present (Offringa et al. 1990). Therefore, the efficiency of gene targeting using HR is very low in plants. The results that are disclosed herein show how to improve the gene targeting efficiency in plants. From the fact that T-DNA integration by IR is strongly reduced in *KU70, RAD50, MRE11, XRS2, LIG4* and *SIR4* deficient yeast strains and T-DNA integration by HR is not affected in such strains, we infer that also in plants, deficient for either of these genes, T-DNA integration by HR is more easily obtained. Recently, we have cloned a *KU70* homologue of *Arabidopsis thaliana* (see figure 2, Bundock 2000, unpublished data). *RAD50, MRE11* and *LIG4* homologues have already been found in *A.thaliana* (GenBank accession numbers AF168748, AJ243822 and AF233527, respectively, see also figure 3, 4 and 5 (Hartung and Puchta 1999). Currently, screenings are being performed to find plants carrying a T-DNA inserted in *AtMRE11, AtKU70* or *AtLIG4.* These knockout plants are used to test whether T-DNA integration by IR is reduced and integration by HR is unaffected, thereby facilitating the detection of T-DNA integration by HR.

The invention will be explained in more detail in the following description, which is not limiting the invention.

EXPERIMENTAL PART

Yeast strains.

The yeast strains that were used are listed in Table 1. Yeast mutants isogenic to the haploid YPH250 strain were constructed using the one-step disruption method (Rothstein 1991). A 1987 bp fragment from the *YKU70* locus was amplified by PCR using the primers hdf1p1 5'-GGGATTGCTTTAAGGTAG-3' and hdf1p2 5'-CAAATACCCTACCCTACC-3'. The PCR product was cloned into pT7Blue (Novagen) to obtain pT7Blue*YKU70*. A 1177 bp *Eco*RV/*Hin*dIII fragment from the *YKU70* ORF was replaced by a 2033 bp *Hin*dIII/*Sma*I *LEU2* containing fragment from pJJ283 (Jones and Prakash 1990), to form pT7Blue*YKU70::LEU2*. In order to obtain YKU70 disruptants Leu$^+$ colonies were selected after transformation of YPH250 with a 2884 bp *Nde*I/*Sma*I fragment from pT7Blue*YKU70::LEU2*. The Expand™ High Fidelity System (Boehringer Mannheim) was used according to the supplied protocol to amplify a 3285 bp fragment from the *LIG4* locus with primers dnl4p1 5'-CGTAAGATTCGCCGAGTATAG-3' and dnl4p2 5'-CGTTTCAAATGGGACCACAGC-3'. The PCR product was cloned into pGEMT (Promega), resulting in pGEMT*LIG4*. A 1326 bp *Bam*HI/*Xho*I fragment from pJJ215 (Jones and Prakash 1990) containing the *HIS3* gene was inserted into the *Bam*HI and *Xho*I sites of pIC20R, resulting in pIC20R*HIS3*. A 782 bp *Eco*RI fragment from the *LIG4* ORF was replaced with a 1367 bp EcoRI *HIS3* containing fragment from pIC20R*HIS3* to construct pGEMT*LIG4::HIS3*. In order to obtain LIG4 disruptants His$^+$ colonies were selected after transformation of YPH250 with a 3854 bp *Nco*I/*Not*I fragment from pGEMT*LIG4::HIS3*. In order to obtain RAD50 disruptants YPH250 was transformed with a *Eco*RI/*Bgl*II fragment from pNKY83 and Ura$^+$ colonies were selected (Alani et al. 1989). A *rad50::hisG* strain was obtained by selecting Ura$^-$ colonies on selective medium containing 5-FOA. Similarly

RAD51 disruptants were obtained were obtained after transformation of YPH250 with a *RAD51::LEU2 XbaI/PstI* fragment from pDG152 and selection of Leu^+ colonies (Schiestl et al. 1994). The *TRP1* marker in pSM21 (Schild et al. 1983) was replaced with a *BglII/XbaI LEU2* containing fragment from pJJ283 (Jones and Prakash, 1990). This resulted in pSM21*LEU2*. Leu^+ RAD52 disruptant colonies were selected after transformation of YPH250 with the *RAD52::LEU2 BamHI* fragment from pSM21*LEU2*. Disruption constructs were transformed to YPH250 by the lithium acetate transformation method as described (Gietz et al. 1992). Disruption of *YKU70, LIG4, RAD50, RAD51* and *RAD52* was confirmed by PCR and Southern blot analysis.

Table 1: Yeast strains

Strain	Genotype	Reference
YPH250	*MATa, ura3-52, lys2-801, ade2-101, trp1-Δ1, his3-Δ200, leu2-Δ1*	Sikorski and Hieter 1989
YPH250*rad51*	*MATa, ura3-52, lys2-801, ade2-101, trp1-Δ1, his3-Δ200, leu2-Δ1, rad51::LEU2*	This study
YPH250*rad52*	*MATa, ura3-52, lys2-801, ade2-101, trp1-Δ1, his3-Δ200, leu2-Δ1, rad52::LEU2*	This study
YPH250*yku70*	*MATa, ura3-52, lys2-801, ade2-101, trp1-Δ1, his3-Δ200, leu2-Δ1, yku70::LEU2*	This study
YPH250*rad50*	*MATa, ura3-52, lys2-801, ade2-101, trp1-Δ1, his3-Δ200, leu2-Δ1,rad50::hisG*	This study
YPH250*lig4*	*MATa, ura3-52, lys2-801, ade2-101, trp1-Δ1, his3-Δ200, leu2-Δ1, lig4::HIS3*	This study
JKM115	*Δho, Δhml::ADE1, MATa, Δhmr::ADE1, ade1, leu2-3,112, lys5, trp1::hisG, ura3-52*	Moore and Haber 1996
JKM129	*Δho, Δhml::ADE1, MATa, Δhmr::ADE1, ade1, leu2-3,112, lys5, trp1::hisG, ura3-52, xrs2::LEU2*	Moore and Haber 1996
JKM138	*Δho, Δhml::ADE1, MATa, Δhmr::ADE1, ade1, leu2-3,112, lys5, trp1::hisG, ura3-52, mre11::hisG*	Moore and Haber 1996
YSL204	*Δho, HMLa, MATa, HMRa, ade1-100, leu2-3,112, lys5, trp1::hisG, ura3-52, hisG'-URA3-hisG', sir4::HIS3*	Lee et al. 1999

Construction of binary vectors.

To construct pSDM8000 a 1513 bp *Pvu*II/*Eco*RV fragment carrying the *KanMX* marker was obtained from pFA6a (Wach et al. 1994) and was ligated into the unique *Hpa*I site of pSDM14 (Offringa 1992). pSDM8001 was made in three cloning steps. A 1476 bp *Bam*HI/*Eco*RI fragment carrying the *KanMX* marker was obtained from pFA6a and ligated into *Bam*HI and *Eco*RI digested pIC20H to form pIC20H*kan*MX. The KanMX marker was inserted between the *PDA1* flanks by replacement of a 2610 bp *Bgl*II fragment from pUC4E1α10 (Steensma et al. 1990) with a 1465 *Bgl*II fragment from pIC20H*kan*MX. A 3721 bp *Xho*I/*Kpn*I fragment from this construct was inserted into the *Xho*I and *Kpn*I sites of pSDM14. The binary vectors pSDM8000 and pSDM8001 were introduced into *Agrobacterium tumefaciens* LBA1119 by electroporation (den Dulk-Ras and Hooykaas 1995).

Cocultivations / T-DNA transfer experiments.

Cocultivations were performed as described earlier with slight modifications (Bundock et al. 1995). *Agrobacterium* was grown overnight in LC medium. The mix of *Agrobacterium* and *S. cerevisiae* cells was incubated for 9 days at 20°C. G418 resistant *S.cerevisiae* strains were selected at 30°C on YPAD medium containing geneticin (200 μg/ml) (Life Technologies/Gibco BRL).

Vectorette PCR.

Chromosomal DNA was isolated using Qiagen's Genomic Tips G/20 per manufacters protocol. 1-2 μg of Genomic DNA was digested with *Eco*RI, *Cla*I, *Pst*I or *Hin*dIII and run on a 1% TBE-gel. Non-radioactive Southern blotting was performed. The membrane was hybridized with a digoxigenine-labeled kanMX probe to determine the size of T-DNA/genomic DNA fragments (*Eco*RI and *Cla*I for RB containing fragments and *Pst*I and *Hin*dIII for LB containing fragments). The kanMX probe, a 792 bp internal fragment of the *KanMX*

marker, was made by PCR using primers kanmxp1 5'-AGACTCACGTTTCGAGGCC-3' and kanmxp2 5'-TCACCGAGGCAGTTCCATAG-3' and a Non-Radioactive DNA Labeling and Detection kit (Boehringer Mannheim). The enzyme showing the smallest band on blot was used for Vectorette PCR, in order to amplify the smallest junction sequence of T-DNA and genomic DNA. Vectorette PCR was performed as described (http://genomewww.stanford.edu/group/botlab/protocols/vectorette.html). The Expand™ High Fidelity System (Boehringer Mannheim) was used to amplify fragments larger than 2.5 kb, whereas sTaq DNA polymerase (SphaeroQ) was used for amplification of fragments smaller than 2.5 kb. Primer kanmxp3 5'-TCGCAGGTCTGCAGCGAGGAGC-3' and kanmxp4 5'-TCGCCTCGACATCATCTGCCCAG-3' were used to amplify RB/genomic DNA and LB/genomic DNA junction sequences, respectively.

T7 DNA Polymerase sequencing.

Vectorette PCR products were cloned in pGEMTEasy (Promega) and sequenced using the T7 polymerase sequencing kit (Pharmacia) according to manufacturers protocol. In order to obtain sequences flanking the RB and LB, primers kanmxp5 5'-TCACATCATGCCCCTGAGCTGC-3' and kanmxp4 were used, respectively.

RESULTS

1. Binary vectors for T-DNA transfer to yeast.

It was previously demonstrated that *Agrobacterium tumefaciens* is able to transfer its T-DNA not only to plants but also to another eukaryote, namely the yeast *Saccharomyces cerevisiae* (Bundock et al. 1995). T-DNA carrying homology with the yeast genome was shown to become integrated by homologous recombination. T-DNA lacking any homology with the yeast genome was integrated randomly into the genome by IR like in plants (Bundock et al. 1995, Bundock and Hooykaas 1996). The T-DNA used in these experiments carried the *S.cerevisiae URA3* gene for selection of Ura$^+$ colonies after T-DNA transfer to the haploid yeast strain RSY12(*URA3Δ*). However, in this system only yeast strains could be used in which the *URA3* gene had been deleted to avoid homology between the incoming T-DNA and the *S.cerevisiae* genome.

We wanted to setup a system in which T-DNA transfer to any yeast strain could be studied. Therefore, two new binary vectors were constructed using the dominant marker *kanMX* (Wach et al. 1994), which confers resistance against geneticin (G418). The T-DNA of pSDM8000 carries only the *KanMX* marker. Since this *KanMX* marker consists of heterologous DNA, lacking any homology with the *S.cerevisiae* genome, we would expect this T-DNA to integrate by IR at a random position in the yeast genome. To be able to compare this with T-DNA integration by homologous recombination pSDM8001 was constructed. The T-DNA of pSDM8001 carries the *KanMX* marker flanked by sequences from the *S.cerevisiae PDA1* locus. The *PDA1* sequences have been shown to mediate the integration of T-DNA by HR at the *PDA1* locus on chromosome V (Bundock et al. 1995).

Cocultivations between *Agrobacterium* strains carrying pSDM8000 and pSDM8001, respectively, and the haploid yeast strains YPH250 and JKM115, respectively, were carried out as described in the experimental part. G418

resistant colonies were obtained at low frequencies for YPH250 (1.6×10^{-7}) and JKM115 (1.2×10^{-5}) after T-DNA transfer from pSDM8000 (Table 2). T-DNA transfer from pSDM8001 generated G418 resistant colonies at higher frequencies (2.4×10^{-5} for YPH250 and 1.8×10^{-4} JKM115, Table 2). The ratio of homologous recombination versus illegitimate recombination is determined by comparing the frequencies of G418 resistant colonies obtained from cocultivations using either pSDM8001 or pSDM8000. This showed that a T-DNA from pSDM8001 was 150-fold more likely to integrate than a T-DNA from pSDM8000 in YPH250 (Table 2). A similar difference was previously seen using T-DNAs with the *URA3* marker (Bundock and Hooykaas 1996). In contrast, T-DNA from pSDM8001 was only 16-fold more likely to integrate than a T-DNA from pSDM8000 in JKM115. There was no significant difference in the frequency of T-DNA transfer to these two yeast strains as was determined by T-DNA transfer experiments in which a T-DNA, that carried the *KanMX* marker and the yeast 2 micron replicon, was employed. Therefore, the differences in the frequencies of T-DNA integration by HR and IR between the yeast strains YPH250 and JKM115, respectively, is most likely contributed to differences in the capacities of their HR and IR recombination machineries.

We confirmed by PCR that the T-DNA from pSDM8001 became integrated at the *PDA1* locus by homologous recombination (data not shown). In order to find out whether the T-DNA from pSDM8000 had integrated randomly by IR yeast target sites for integration were determined from 8 G418 resistant YPH250 colonies by Vectorette PCR (for detailed description see materials and methods). Chromosomal DNA was isolated and digested with a restriction enzyme that cuts within the T-DNA. A Vectorette was ligated to the digested DNA and a PCR was performed using a T-DNA specific and a Vectorette specific primer. The PCR product obtained was cloned into pGEMTEasy and sequenced using a T-DNA specific primer. The position of the T-DNA insertion was determined by basic BLAST search of the yeast genome (http://www-genome.stanford.edu/SGD). We were thus able to map the position

of the T-DNA insertions of all 8 G418 resistant colonies analyzed. They were present at different positions spread out over the genome. Comparison of the T-DNA sequence and yeast target site sequences did not reveal any obvious homology. These data show that the T-DNA from pSDM8000 had integrated via an IR mechanism as expected.

The following characteristics have previously been observed for T-DNAs integrated by IR: a) the 3' end of the T-DNA is usually less conserved compared to the 5' end, b) microhomology is sometimes present between the T-DNA ends and the target site, c) integration is often accompanied by small deletions of the target site DNA (Matsumoto et al. 1990, Gheysen et al. 1991, Mayerhofer et al. 1991, Bundock and Hooykaas 1996). Similar characteristics were seen in the currently analyzed 8 T-DNA insertions. In 3 strains we observed microhomology of 2 – 6 bp between the LB and yeast target site (figure 1, WT.51 was taken as an example). In 5 strains deletions of 1 – 5 bp of yeast target site DNA was found and we observed deletions, varying from 1 – 112 bp, of the 3' end of the T-DNA in 7 out of 8 analyzed strains. In only 1 strain the 3' end appeared to be intact. The 5' end of the T-DNA was conserved in almost all strains. In only 2 strains we observed small deletions of 1 and 2 bp at the 5' end of the T-DNA.

Thus, we can conclude that the T-DNA from pSDM8000 had integrated via the same IR mechanism described before.

Table 2: frequencies of T-DNA integration by IR relative to integration by HR in recombination defective yeast strains

Strain	Genotype	Freq. of IR[a]	Freq. of HR	Absolute IR/HR ratio[b]	Standardized IR/HR ratio[c]
YPH250	WT	1.6×10^{-7}	2.4×10^{-5}	0.007	1
YPH250 *rad51*	*rad51Δ*	1.4×10^{-7}	1.5×10^{-6}	0.09	14
YPH250 *rad52*	*rad52Δ*	3.8×10^{-7}	2.5×10^{-6}	0.15	23
YPH250 *yku70*	*yku70Δ*	$<3.2 \times 10^{-9}$	3.3×10^{-5}	<0.0001	<0.01
YPH250 *rad50*	*rad50Δ*	8.0×10^{-9}	3.5×10^{-5}	0.0002	0.03
YPH250 *lig4*	*lig4Δ*	3.7×10^{-9}	2.3×10^{-5}	0.0002	0.02
JKM115	WT	1.2×10^{-5}	1.8×10^{-4}	0.07	1
JKM129	*xrs2Δ*	2.7×10^{-7}	5.1×10^{-5}	0.005	0.08
JKM138	*mre11Δ*	2.9×10^{-7}	7.5×10^{-5}	0.004	0.06
YSL204	*sir4Δ*	1.5×10^{-7}	1.8×10^{-5}	0.008	0.13

[a] Averages of 2 or more independant experiments are shown. Frequencies are depicted as the number of G418 resistant colonies devided by the output number of yeast cells (cells/ml).

[b] The frequency of T-DNA integration by IR (pSDM8000) devided by the frequency of T-DNA integration by HR (pSDM8001).

[c] The ratio of IR/HR in the mutant strain devided by the ratio of IR/HR in the wildtype strain.

2. Host-specific factors involved in random T-DNA integration.

The observation that the T-DNA from pSDM8000 integrates by IR into the yeast genome allowed us to use this system to study the effect of host factors on the process of integration. Many proteins involved in various forms of DNA recombination have been identified in yeast. In order to determine the roles of a representative set of these enzymes in T-DNA integration, we compared T-DNA transfer and integration in wildtype yeasts with that of strains carrying disruptions of the genes encoding several recombination proteins. The *RAD51, RAD52, KU70, RAD50* and *LIG4* genes were deleted from YPH250 using the one step disruption method. Yeast strains carrying deletions in *MRE11, XRS2* and *SIR4* in the JKM115 background were kindly provided by Dr. J. Haber. The results of cocultivations with these yeast strains are shown in Table 2.

In *rad51* and *rad52* mutants, which are impaired in homologous recombination, the frequency of T-DNA integration by HR was 16- and 9-fold lower, respectively, than observed for the wildtype (Table 2). This showed that *RAD51* and *RAD52* play a role in T-DNA integration by homologous recombination. In the IR defective *ku70, rad50, lig4, mre11, xrs2* and *sir4* mutants the frequency of T-DNA integration by HR did not differ significantly from that observed for wildtype (Table 2). This showed that these genes do not play a role in T-DNA integration by homologous recombination.

The frequency of T-DNA integration by IR in a *rad51* mutant did not differ significantly from that observed for wildtype, whereas in a *rad52* mutant the frequency was about 2-fold higher (Table 2). This showed that *RAD51* and *RAD52* are not involved in T-DNA integration by IR. The product of the *RAD52* gene may compete with IR-enzymes for the T-DNA and thereby inhibits integration by IR to some extent. Strikingly, in *rad50, mre11, xrs2, lig4* and *sir4* mutants the frequency of T-DNA integration by IR was reduced dramatically: 20- to more than 40-fold (Table 2). T-DNA integration by IR seemed to be completely abolished in the *ku70* mutant. We did not obtain any

G418 resistant colonies from several cocultivation experiments. This strongly suggests that KU70 plays an important role in random T-DNA integration in yeast.

Since T-DNA integration by HR is normal in these mutants, these results clearly show that the yeast genes *KU70, RAD50, MRE11, XRS2, LIG4* and *SIR4* are involved in T-DNA integration by illegitimate recombination.

3. Chromosomal distribution of integrated T-DNA copies in IR defective *S.cerevisiae.*

From several cocultivation experiments with the *rad50, mre11, xrs2, lig4* and *sir4* mutants we obtained a small number of G418 resistant colonies. The T-DNA structure was determined for a number of these lines. To this end chromosomal DNA was isolated from these G418 resistant colonies and subjected to vectorette PCR to amplify junction sequences of genomic DNA and T-DNA. PCR products were cloned and sequenced. The yeast sequences linked to the T-DNA were used in a BLAST search at http://www-genome.stanford.edu/SGD to determine the T-DNA integration sites.

Strikingly, analysis of LB/genomic DNA junctions revealed that in 2 out of 3 *rad50,* 4 out of 6 *mre11* and 2 *xrs2* strains analyzed, T-DNAs had integrated in telomeres or subtelomeric regions (*rad50k.1, rad50k.6, mre11k.8, mre11k.11, mre11k.14, mre11k.17, xrs2k.1* and *xrs2k.17*; Table 3 and figure 1). *S. cerevisiae* telomeres generally consist of one or more copies of the Y' element followed by telomerase-generated C(1-3)A/TG(1-3) repeats (Zakian 1996). In 2 *rad50* strains, 2 *mre11* strains and 1 *xrs2* strain the LB was found to be fused to this typical telomerase-generated C(1-3)A/TG(1-3) repeat (*rad50k.1, rad50k.6, mre11k.14, mre11k.17* and *xrs2k.1*; figure 1). Besides, we also found one T-DNA insertion in a Ty LTR element in the *mre11* mutant and 2 insertions in the rDNA region, present in chromosome XII, in the *mre11* and *rad50* mutants (*mre11k.5, mre11k.4* and *rad50k.5,* respectively; Table 3 and figure 1).

The 3' end of the T-DNA was truncated in all strains. Deletions of 3 – 11 bp of the 3'end of the T-DNA were observed (figure 1). Microhomology between the 3' end of the T-DNA and yeast target site was only found in 2 lines (5 bp in *mre11k.4* and 4 bp in *mre11k.14*; figure 1). For the T-DNA copies present at the yeast telomeres, the RB/genomic DNA junction sequences could not be obtained from these strains using vectorette PCR. This was only possible for the *rad50* and *mre11* strains carrying the T-DNA in the rDNA region on chromosome XII. In both strains the RB was intact and no homology between the 5' end of the T-DNA and the yeast target site was found (data not shown in figure 1).

Previously, target sites for T-DNA integration in the genome of *S.cerevisiae* strain RSY12 were determined (Bundock and Hooykaas 1996, Bundock 1999). In 4 out of 44 strains analyzed, T-DNA copies were integrated in rDNA, Ty LTR elements (in 2 strains) and a subtelomeric located Y' element, respectively. In addition, we determined the position of T-DNA integration in ten *S.cerevisiae* YPH250 strains. We did not find any T-DNA insertions in rDNA, LTR elements or subtelomeric/telomeric regions amongst these ten. Pooling all insertions analyzed in wildtype (54), in 2 out of 54 strains analyzed (4%) insertions were found in a Ty LTR element and in two other strains in the rDNA repeat (2%) and a subtelomeric region (2%), respectively. In contrast, we report here that T-DNA in yeast strains mutated in *RAD50, MRE11* or *XRS2* T-DNA integrates preferentially in (sub)telomeric regions (8 out of 11 lines: ~73%) of *rad50, mre11* and *xrs2* mutants (table 3). From the remaining strains two T-DNAs were present in rDNA and one in a Ty LTR element, respectively. Apparently, the rDNA repeat is also a preferred integration site in these mutants (~18% vs. ~2% in the wildtype).

Telomeres consist of a large array of telomerase-generated C(1-3)A/TG(1-3) repeats (~350 bp). In the subtelomeric regions two common classes of Y' elements, 6.7 and 5.2 kb, can be found (in most strains chromosome I does not contain Y') (Zakian and Blanton 1988), making the average size of these

regions ~6,0 kb. Thus, the yeast genome contains (16 x 2 x 0.35) + (15 x 2 x 6,0) = 191 kb of subtelomeric/telomeric sequences. The yeast genome is 12,052 kb in size, which means that only 1.6% of the genome consists of subtelomeric/telomeric sequences. In accordance with this, we observed in only 2% of the wildtype strains T-DNA copies inserted in a subtelomeric region, which we would expect on the basis of random T-DNA integration. In contrast, in the *rad50, mre11* and *xrs2* mutants 73% of the T-DNA insertions were found in the (sub)telomeric region.

Analysis of 7 lines revealed that in the *sir4* mutant T-DNA was integrated randomly into the yeast genome. So, although *SIR4* has an effect on the efficiency of T-DNA integration by IR, the pattern of T-DNA distribution in the transformants seems similar as in the wildtype strain. In the *sir4* mutant T-DNA integration by IR was characterized by truncation of the 3' end of the T-DNA, deletions at the target site and microhomology between the LB and the target site (data not shown), like this was observed for T-DNA integration by IR in the wildtype.

These results clearly show that in the *rad50, mre11* and *xrs2* mutants the T-DNA, if integrated at all, becomes preferentially inserted in telomeres or subtelomeric regions and that the genomic distribution of integrated T-DNAs is altered when compared to wildtype. However, disruption of *SIR4* did affect the efficiency of T-DNA integration by IR, but not the characteristics of this process.

Table 3: genomic distribution of T-DNA integrated by IR in *rad50, mre11* and *xrs2* mutants in comparison with the wildtype after T-DNA transfer from pSDM8000

Yeast strain	*(Sub)Telomeric region*	*LTR*	*rDNA*	Elsewhere
rad50 mutant	2	0	1	0
mre11 mutant	4	1	1	0
xrs2 mutant	2	0	0	0
wildtype	1	2	1	50

DESCRIPTION OF FIGURES

Figure 1: Junction sequences of T-DNA and *S.cerevisae* genomic DNA. *S.cerevisiae* YPH250 (WT), *rad50, mre11* and *xrs2* strains were cocultivated with LBA1119(pSDM8000). G418 resistant colonies were obtained. Chromosomal DNA was isolated and subjected to Vectorette PCR to determine the sequence of genomic DNA flanking the T-DNA. Position of T-DNA integration was determined by basic BLAST search of the yeast genome at http:/www.genome-stanford.edu/SGD. The Watson strand of genomic DNA that is fused to the LB or RB is shown in italics. Bold sequences represent sequence homology between the LB and target site. The filler DNA sequence is underlined and depicted in italics. The numbers above the LB sequences represents the number of bp deleted from the LB. Tel. = telomeric, Subtel. = subtelomeric and Int. = intergenic.

Figure 2: Alignment of KU70 homologues. Sc = *Saccharomyces cerevisiae,* Hs = *Homo sapiens* and At = *Arabidopsis thaliana.* Perfect identity is depicted as black boxes, homology is depicted as grey boxes and dashes were used to optimise alignment.

Figure 3: Alignment of LIG4 homologues. Sc = *Saccharomyces cerevisiae,* Hs = *Homo sapiens* and At = *Arabidopsis thaliana.* Perfect identity is depicted as black boxes, homology is depicted as grey boxes and dashes were used to optimise alignment.

Figure 4: Alignment of MRE11 homologues. Sc = *Saccharomyces cerevisiae*, Hs = *Homo sapiens* and At = *Arabidopsis thaliana*. Perfect identity is depicted as black boxes, homology is depicted as grey boxes and dashes were used to optimise alignment.

Figure 5: Alignment of RAD50 homologues. Sc = *Saccharomyces cerevisiae*, Hs = *Homo sapiens* and At = *Arabidopsis thaliana*. Perfect identity is depicted as black boxes, homology is depicted as grey boxes and dashes were used to optimise alignment.

REFERENCE LIST

Alani, E., Subbiah, S., and Kleckner, N. 1989. The yeast *RAD50* gene encodes a predicted 153-kD protein containing a purine nucleotide-binding domain and two large heptad-repeat regions. Genetics 122:47-57.

Bundock, P. 1999. *Agrobacterium tumefaciens*-mediated transformation of yeast and fungi. Thesis Leiden University.

Bundock, P., den Dulk-Ras, A., Beijersbergen, A., and Hooykaas, P. J. J. 1995. Trans-kingdom T-DNA transfer from *Agrobacterium tumefaciens* to *Saccharomyces cerevisiae*. EMBO J. 14:3206-3214.

Bundock, P. and Hooykaas, P. J. J. 1996. Integration of *Agrobacterium tumefaciens* T-DNA in the *Saccharomyces cerevisiae* genome by illegitimate recombination. Proc. Natl. Acad. Sci U.S.A 93:15272-15275.

deGroot, M. J. A., Bundock, P., Hooykaas, P. J. J., and Beijersbergen, A. G. M. 1998. *Agrobacterium tumefaciens*-mediated transformation of filamentous fungi. Nat. Biotech. 16:839-842.

den Dulk-Ras, A. and Hooykaas, P. J. J. 1995. Electroporation of *Agrobacterium tumefaciens*. Methods Mol. Biol. 55:63-72.

Gheysen, G., Villarroel, R., and Van Montagu, M. 1991. Illegitimate recombination in plants: a model for T-DNA integration. Genes Dev. 5:287-297.

Gietz, D., St.Jean, A., Woods, R. A., and Schiestl, R. H. 1992. Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res. 20:1425.

Hansen, G. and Chilton, M. D. 1999. Lessons in gene transfer to plants by a gifted microbe. Curr. Top. Microbiol. Immunol. 240:21-57.

Hartung, F. and Puchta, H. 1999. Isolation of the complete cDNA of the *MRE11* homologue of *Arabidopsis* (Accession No. AJ243822) indicates conservation of DNA recombination mechanisms between plants and other eukaryotes. Plant Phys. 121:312-312.

Hooykaas, P. J. J. and Beijersbergen, A. G. M. 1994. The virulence system of *Agrobacterium tumefaciens*. Annu. Rev. of Phytopathol. 32:157-179.

Jones, J. S. and Prakash, L. 1990. Yeast *Saccharomyces cerevisiae* selectable markers in pUC18 polylinkers. Yeast 6:363-366.

Lee, S. E., Paques, F., Sylvan, J., and Haber, J. E. 1999. Role of yeast *SIR* genes and mating type in directing DNA double-strand breaks to homologous and non-homologous repair paths. Curr. Biol. 9:767-770.

Matsumoto, S., Ito, Y., Hosoi, T., Takahashi, Y., and Machida, Y. 1990. Integration of Agrobacterium T-DNA into a tobacco chromosome: possible involvement of DNA homology between T-DNA and plant DNA. Mol. Gen. Genet. 224:309-316.

Mayerhofer, R., Koncz-Kalman, Z., Nawrath, C., Bakkeren, G., Crameri, A., Angelis, K., Redei, G. P., Schell, J., Hohn, B., and Koncz, C. 1991. T-DNA

integration: a mode of illegitimate recombination in plants. EMBO J. 10:687-704.

Moore, J. K. and Haber, J. E. 1996. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. Mol. Cell. Biol. 16:2164-2173.

Offringa, R. 1992. Gene targeting in plants using the *Agrobacterium* vector system. Thesis Leiden University.

Offringa, R., de Groot, M. J., Haagsman, H. J., Does, M. P., van den Elzen, P. J., and Hooykaas, P. J. J. 1990. Extrachromosomal homologous recombination and gene targeting in plant cells after *Agrobacterium* mediated transformation. EMBO J. 9:3077-3084.

Rothstein, R. 1991. Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Methods Enzymol. 194:281-301.

Schiestl, R. H., Zhu, J., and Petes, T. D. 1994. Effect of mutations in genes affecting homologous recombination on restriction enzyme-mediated and illegitimate recombination in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 14:4493-4500.

Schild, D., I. L. Calderon, C. R. Contopoulou, and R. K. Mortimer. 1983. Cloning of yeast recombination repair genes and evidence that several are nonessential genes, p. 417-427. *In*: E. C. Friedberg and B. A. Bridges (eds.), Cellular responses to DNA damage. Alan R. Liss Inc., New York.

Sikorski, R. S. and Hieter, P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae.* Genetics 122:19-27.

Steensma, H. Y., Holterman, L., Dekker, I., van Sluis, C. A., and Wenzel, T. J. 1990. Molecular cloning of the gene for the E1 alpha subunit of the pyruvate dehydrogenase complex from *Saccharomyces cerevisiae.* Eur. J. Biochem. 191:769-774.

Vergunst, A. C., Schrammeijer, B., den Dulk-Ras, A., de Vlaam, C. M. T., Regensburg-Tuïnk, T. J. G., and Hooykaas, P. J. J. 2000, VirB/D4 dependent protein translocation from *Agrobacterium* into plant cells. Science: in press.

Wach, A., Brachat, A., Pohlmann, R., and Philippsen, P. 1994. New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae.* Yeast 10:1793-1808.

Zakian, V. A. 1996. Structure, function, and replication of *Saccharomyces cerevisiae* telomeres. Annu. Rev. Genet. 30:141-172.

Zakian, V. A. and Blanton, H. M. 1988. Distribution of telomere-associated sequences on natural chromosomes in *Saccharomyces cerevisiae.* Mol. Cell. Biol. 8:2257-2260.

Zrenner, R., Willmitzer, L., and Sonnewald, U. 1993. Analysis of the expression of potato uridinediphosphate-glucose pyrophosphorylase and its inhibition by antisense RNA. Planta, 190:247-252

Claims

1. A method to direct integration of a nucleic acid of interest to a pre-determined site, whereby said nucleic acid has homology at or around the said pre-determined site, in an eukaryote with a preference for non-homologous recombination, comprising steering an integration pathway towards homologous recombination.

2. A method to direct nucleic acid integration according to claim 1, comprising providing a mutant of a component involved in non-homologous recombination.

3. A method to direct nucleic acid integration according to claim 1 or 2, comprising inhibiting a component involved in non-homologous recombination.

4. A method according to claim 2 or 3 wherein said component involved in non-homologous recombination comprises *ku70*, *rad50*, *mre11*, *xrs2*, *lig4* or *sir4*.

5. A method to direct integration of a nucleic acid of interest to a sub-telomeric and/or telomeric region in an eukaryote with a preference for non-homologous recombination by providing a mutant of a component involved in non-homologous recombination.

6. A method to direct integration of a nucleic acid of interest to a sub-telomeric and/or telomeric region in an eukaryote with a preference for non-homologous recombination, comprising inhibiting a component involved in non-homologous recombination

7. A method to direct integration according to claim 5 or 6 wherein said component involved in non-homologous recombination comprises *rad50*, *mre11* or *xrs2*.

8. A method according to anyone of claims 1 to 7 wherein said eukaryote comprises yeast.

9. A method according to anyone of claims 1-8 comprising transiently inhibiting integration via non-homologous recombination.

10. A method according claim 9 wherein said transiently inhibiting is provided by an *Agrobacterium* Vir-fusion protein capable of inhibiting a component involved in non-homologous recombination.

11. A method to direct nucleic acid integration according to claim 10 wherein said *Agrobacterium* Vir fusion protein comprises VirF or VirE2.

12. A method according to claim 10 or 11 wherein said component involved in non-homologous recombination comprises *ku70, rad50, mre11, xrs2, lig4* or *sir4.*

13. A method according to anyone of the aforegoing claims wherein said nucleic acid of interest comprises an inactive gene to replace an active gene.

14. A method according to anyone of claims 1-12, wherein said nucleic acid of interest comprises an active gene to replace an inactive gene.

15. A method according to anyone of claims 1-12, wherein said nucleic acid of interest encodes a therapeutic proteinaceous substance.

16. A method according to anyone of claims 1-12, wherein said nucleic acid of interest encodes a substance conferring resistance for an antibiotic substance to a cell.

17. A method according to anyone of claims 1-12, wherein said nucleic acid of interest confers a desired property to said eukaryotic cell.

18. A method according to anyone of the aforegoing claims wherein said nucleic acid of interest is part of a gene delivery vehicle.

19. Use of a method according to anyone of claims 1 to 18 for improvement of gene targeting efficiency.

EPO - DG 1

22. 12. 2000

60

Title: Nucleic acid integration in eukaryotes

Abstract

The invention relate to the field of molecular biology and cell biology. It particularly relates to methods to direct integration of a nucleic acid of interest towards homologous recombination and uses thereof. The present invention discloses factors involved in integration of a nucleic acid by illegitimate recombination which provides a method to direct integration of a nucleic acid of interest to a pre-determined site, whereby said nucleic acid has homology at or around the said pre-determined site, in an eukaryote with a preference for non-homologous recombination comprising steering an integration pathway towards homologous recombination. Furthermore, the invention provides a method to direct integration of a nucleic acid of interest to a sub-telomeric and/or telomeric region in an eukaryote with a preference for non-homologous recombination.

FIGURE 1

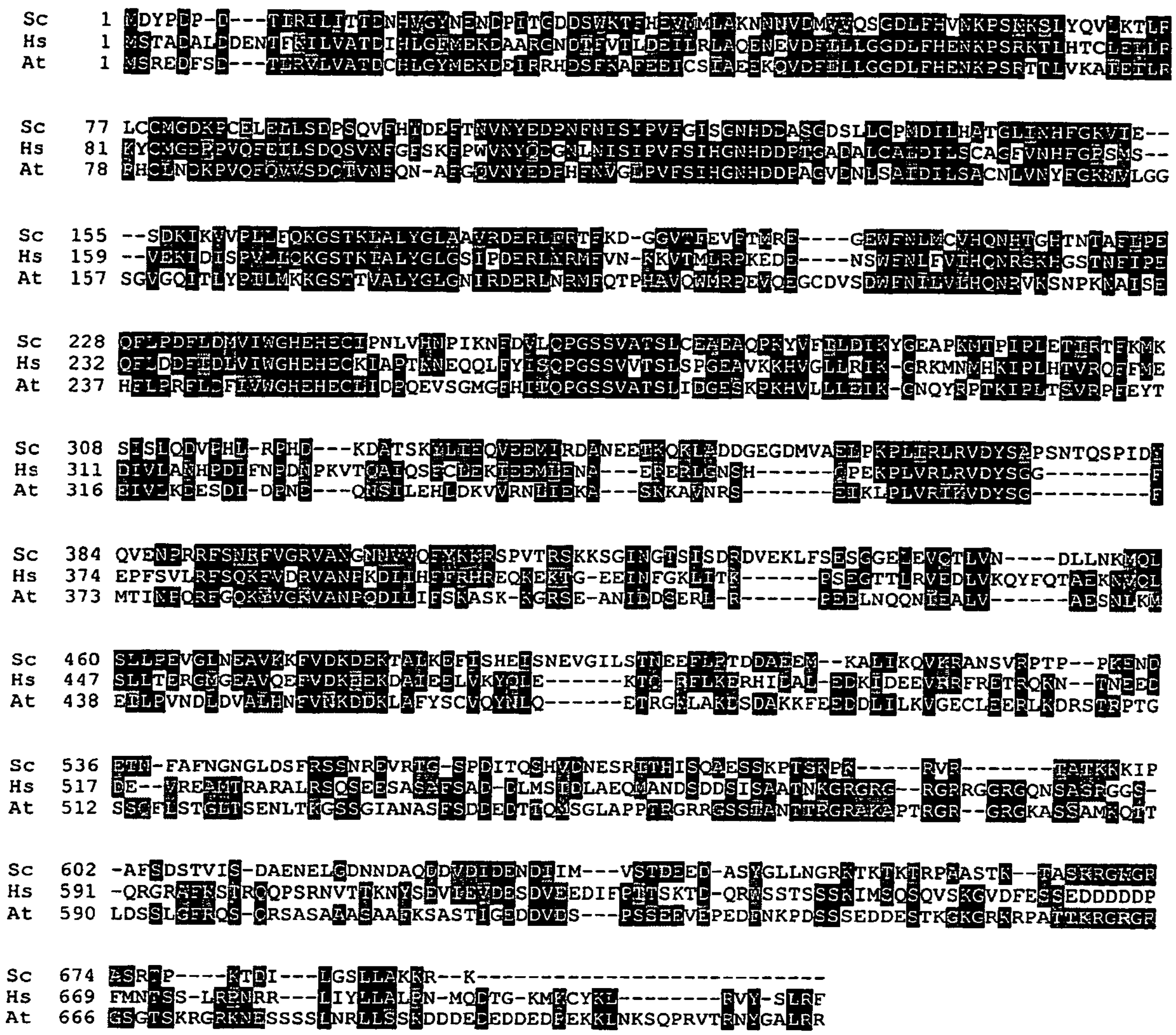
Strain	LB' T-DNA RB' CAGGATATATTCAATTGTAAAT-CTC---CGA-GG	Chromosome, coordinate and location
WT.51	5' *ATTGTATT*ATATATTCAATTGTAAAT-CTC---CGA-GG*TA* 3' (-4)	XIV, 185311 (1 bp of target site DNA deleted), int. region
rad50k.1	5'*TGTGGGTGT*GATATTCAATTGTAAAT-CTC—CGA-GG 3' (-6)	XV, 1091277, tel. region
rad50k.5	5' *GGGGGCATCAG*TATTCAATTGTAAAT-CTC---CGA-GG 3' (-7)	XII, 465986, rDNA region
rad50k.6	5' *GAGGTAGATGTGAGAGAGTGTGTGTGGGTGT*GAAGTCGA 3' (-25)	XV, 1091276, tel. region
mre11k.4	5' *TCTGGTA*GATATATTCAATTGTAAAT-CTC---CGA-GG 3' (-3)	XII, 459692/468829, rDNA region
mre11k.5	5' *CACATATTTCTC*ATTCAATTGTAAAT-CTC---CGA-GG 3' (-8)	VII/X/XIII, 536090 OR 541678/ 472487 OR 483659/196667, LTR
mre11k.8	5' *CGACTACTTTATATCC*AATTGTAAAT-CTC---CGA-GG 3' (-11)	XIV, 6060, subtel. region
mre11k.11	5' *GAAGAACCCATT*ATTCAATTGTAAAT-CTC---CGA-GG 3' (-7)	XIV, 4866, subtel. region
mre11k.14	5' *TGGGTGTGGG*TTATTCAATTGTAAAT-CTC---CGA-GG 3' (-7)	VIII, 562588, tel. region
mre11k.17	5' *TGGGTGTGGTGTGT*TCAATTGTAAAT-CTC---CGA-GG 3' (-9)	XII, 5727, subtel. region
xrs2k.1	5' *TGTGTGGGTGTGGG*TCAATTGTAAAT-CTC---CGA-GG 3' (-10)	IX/X, 69/52, tel. region
xrs2k.17	5' *CGTCA*AGGATATATTCAATTGTAAAT-CTC---CGA-GG 3' (-1)	XII, 1071797, subtel region

```
Sc    1 ------MRSVTNAFGNSGELNDQVDETGYRKEDIHEGEFFCIELSETMFKESSDLEYKSELLEILESEDEEMSQLWITRP
Hs    1 MSGWESYYKTEGDEEAEEEQEENLEASGDYKYSGRDSLIFLVDASKAMFESQSEDELT-PFDMSIQCIQSVYISKIISSD
At    1 --------------------------------------------------------------------------------

Sc   75 GTAIGCYFYYCNREDAKEGIYELEPLRDINATFMKKLNDLLEDLSSGRISLYDLFMFQQTGSEKQVRLSVLFTFMLDTFL
Hs   80 RDLLAVVFYGTEKDKNSVNFKNIYVLQELDNPGAKRI--LELDQFKGQQGQKRFQDMMGHGSDYSL-SEVLW--VCANLF
At    1 -----------------------------------------------------------ENSL-YSALW--VAQALI

Sc  155 EEIPGQKQLSNKRVFLFTEIDKPQEAQD-IDERARLRR---LTIDLFENKVNFATFFIGYADKPFDN-EFYSDILQLGSH
Hs  155 SDV--GFKMSHKRIMLFTNEDNPHGND--SAKASRART---KAGDLRDTGIFLDLMHLKKPGG-FDISLFYRDIISIAE-
At   16 RKG--SLKTADKRMFLFTNEDDPFGSMRISVKEDMTRTTLQRAKDAQDLGISIELLPLSQPDKQFNITLFYKDLIGLNS-

Sc  230 TNENTGLDSEFDGPSTKPIDAKYIKSRILRRKEVKRIMFQCPLILDEKTNFIVGVKGYTMYTHEKAGVRYKLVYEHEDIR
Hs  226 -DE----DLRVHFEESSKLED--LLRKVRAKETRKRALSRLKLKLNKD--IVISVGIYNLVQKALKPP---PIKLYRETN
At   93 -DE----LTEEMPSVGQKLED--MKDQLKKRVLAKRIAKRITFVICDG--LSIELNGYALLRPAIPGS---ITWLDSTTN

Sc  310 QEAYSKRKFLNPITG-EDVTGKTVKVYPYGDLDINLSDSQDQIVMEAYTQKDAFLKIIGFRSSSKSIHYFNNIDKSSFIV
Hs  294 EPVKTKTRTFNTSTGGLLLPSDTKRSQIYGSRQIILEKEETEELKRFD---DPGLMLMGFKPLV-LLKKHHYLRPSLFVY
At  161 LPVKVERSYICTDTG-AIMQDPIQRIQPYKNQNIMFTVEELSQVKRIS---TGHLRLLGFKPLS-CLKDYHNLRPSTFLY

Sc  389 PDEAKYEGSIRTLASLLKILRKKDKIAILWGKLKSNSHPSLYTLSPSS--VKDYN-----EGFYLYRVPFLDEIRKFPSL
Hs  370 PEESLVIGSSTLFSALLIKCLEKEVAALCRYTPRRNIPPYFVALVPQEEELDDQKIQVTPPGFQLVFLPFADDKRKM---
At  236 PSDKEVIGSTRAFIALHRSMIQLERFAVAFYG--GTTPPRLVALVAQD-ETESDGGQVEPPGIHMIYLPYANDIRDIDEL

Sc  462 LSYDDGSEHKEDYDNMKRVTQSIMGYFNLRDGYNPSDFKNPLLQKHYKVLHDYLL---QIETTFDENETPNTKKDR--MM
Hs  447 ----PFTEKIMATPEQVGKMKAIVEKLRFT--YRSDSFENPVLQQHFRNLEALALDLMEPEQAVDLTLPKVEAMNK-RLG
At  313 HSK-PGVAXPRASDDQLKKASALMRRLELKD-FSVCQFANPALQRHYAILQAIALDENELRETRDETLPDEEGMNRPAVV

Sc  537 REDESLRKLYVIRNKILESE-------KSEDPIIQRLNKYVKIWN---------MFYKKFN---------DDNISIKEEKK
Hs  520 SLVDEFKELVYPPDYNPEGK-VTKRKHDNEGSGSKRP-KVEYSEEELKTHISKGTLGKFTVPMLKEACRAYGLKSGLKKQ
At  391 KAIEQFKQSIYGDDPDEESDSGAKEKSKKRKAGDADDGKYDYIEL-AKT----GKLKDLTVVELKTYLTANNLLVSGKKE

Sc  593 PFDKRPKFNI--
Hs  598 ELLEALTKHFQD
At  466 VLINRILTHIGK
```

FIGURE 3


```
Sc     1 MISALDSIPEPQNFAPSPDFKWLCEELLFVKIHEVQINGTAGTGKSRSFKYYEIISNFVEMWRKTVGNNLYPALVLALPYR
Hs     1 -------------------------------------------------------------------------------
At     1 -------MTEEIKKS-------RLVSLFNWIQKSKTSSQKRSKFRKFLDTYCKPSDYFVAMR------LIIPSLD-----R

Sc    81 DFRIYNIKDYVLIRTICSYLKLPKNSATEQRLKDWK--QRVGKGGNLS--SLLVEEIAKRRAEPSSKEITIDNVNHYLDS
Hs    11 ERMAYGIKETMLAKLYIELLNLPRDGKDALKLLNYR--TPTGTHGDAGDFAMIAYFVLKPR-CLQKGSLTIQQVNDLLDS
At    57 ERGSYGLKESVLATCLIEALGISRDAPDAVRLLNWRKGGTAKAGANAGNFSLIAAEVLQRRQGMASGGLTIKELNDLLDR

Sc   157 LSGDRFASGRGFKSLVKSKPFLHCVENMSFVELKYFFDIVLKHRVLGGQEHKLLHCWHPDAQDYLSVISDLKMVTSKLYD
Hs    88 IASN---NSAKRKDLIKKS-LLQLITQSSALEQKWLIRMIIKDLKLGVSQQTIFSVFHNDAAELHNVTTDLEKVCRQLHD
At   137 LASS---ENRAEKTLVLSH-LIQKTN---AQEMKWVIRIILKDLKLGMSEKSIFQEFHPDAEDLFNVTCDLKLVCEKLRD

Sc   237 PEVRLKDDDLSIKVGFAFAPQLAKKVNLSYEKICRTLHDDFLVEEKMDGERIQVHYMNYGESIKEFSRRGIDYTYLYGAS
Hs   164 PSVGLSD--LSITLFSASKPMLAA-IADIEHIEKDMKHQSFYLETKLDGERMQMHKD--GDVYKYFSRNGYNYTDQFGAS
At   210 RHQRHKR--QDIEVGKAVRPQLAMRIGDVNAAWKKLHGKEVVAECKFDGDRIQLHKN--GTDIHYFSRNFLDHSEYAHAM

Sc   317 LSSGTLTG--QHLRFTDSVKECVLDGEMVTFDAKRRVILPFGLVKGSAKEALSFNSINNVDFHPLYMVFDLLYLNGTSLTP
Hs   239 PTEGSLTPFIHNAFKADIQICILDGEMMAYNPNTQTFMQKGTKFDIK---R---MVEDSDLQTCYCVFDVLMVNNKKLGH
At   286 SDLIVQN------ILVE--KCILDGEMLVWDTSLNRFAEFGSNQEIAKAAR---EGLDSHKQLCYVAFDVLYVGDTSVIH

Sc   395 LPLHQRKQYLNSILSPLKNIVEIVRS-----------------SRCYGVESIKKSLEVAISLGSEGVVLKYYNSSYNVASR
Hs   313 ETLRKRYEILSSIFTPIPGRIEIVQK-----------------TQAHTKNEVIDALNEAIDKREEGIMVKQPLSIYKPDKR
At   355 QSLKERHELLKKVVKPLKGRLEVLVPEGGLNVHRPSGEPSWSIVVHAAADVERFFKETVENRDEGIVLKDLESKWEPGDR

Sc   459 NNNWLKVKPEYLEEFGENLDLIVIGRDSGKKDSFMLGLLVLDEEEYKKHQGDSSEIVDHSSQEKHIQNSRRVKKIESFC
Hs   377 GEGWLKIKPEYVSGLMDELDILIVGGYWGKGS------------------RGGMMSHFLCAVAEKPPPGEKPSV--FHTLS
At   435 SGKWMKLKPEYIR-ACADLDVLIIGGYYGSGR------------------RGGEVAQFLVALADRAEANVYPRR--FMSFC

Sc   539 SLANGLSQEEFKEIDRKTRGHWKR-TSEVAPPASILEFG---SKIPAEWID-PSESIVLEIKSRSLDNTETNMQKYATNC
Hs   438 RVGSGCTMKELYDLGLKLAKYWKP-FHRKAPPSSILCGT----EKPEVWIE-PCNSVIVQIK----AAEIVPSDMYKTGC
At   495 RVGTGLSDDELNTVVSKLKPYFRKNEHPKKAPPSFYQVTNHSKERPDVWIDSPEKSIILSITS---DIRTIRSEVFVAPY

Sc   614 TLYGGYCKRIRYDKEWTDCYTLNDLYESRTVKSNPSYQAERSQLG------LIRKKRKRVLISDSFHQNRKCLPISNIFA
Hs   508 TLRFPRIEKIRDDKEWHECMTLDDLEQLRGKASGKLASKHLYIGGDDEPQEKKRKAAPKMKKVIGIIEHLKAPNLTNVNK
At   572 SLRFPRIDKVRYDKPWHECLDVQAFVELVNSSNGTTQKQKESESTQDNPKVNKSSKRGEKKNVSLVPSQFIQTDVSDIKG

Sc   688 GLLFEVLSDEVTEDTGIRIEPAELEKTIVEHGGKLIYNVILKRHSIGDVRLISCKTTTECKALIDRG--YDILHPNWVLD
Hs   588 ISNIFEDVEFCVMSGTDSQPKPDLENRIAEFGGYIVQNPG------PDTYCVIAGSENIRVKNIILSNKHDVVKPAWLLE
At   652 KTSIFSNMIFYFVNVPRSHSLETFHKMVVENGGKFSMNLNN-----SVTHCIAAESSGIKYQAAKRQ--RDVIHFSWVLD

Sc   766 CIAYKRLLLIEPNYCFNVSQKMRAVAEKRVDCLGDSFENDISETKLSSLYKSQLSLPPMGELEIDSEVRRFPLFLFSNR-
Hs   662 CFKTKSFVPWQPRFMIHMCPSTKEHFAREYDCYGDSYFIDTDLNQLKEVFSGIKNSNEQTPEEMASLIADLEYRYSWDCS
At   725 CCSRNKMLPLLPKYFLHLTDASRTKLQDDIDEFSDSYYWDLEDLEGLKQVLSNAKQS--EDSKSIDYYKKKLCPEKRWSC-

Sc   845 ---EAYVPRRKESTEDDIIEKKIKLFGGKITDQQSLCNLIIIPYTDPILP---RDCKNEVHEKIKEQIKASDTI------
Hs   742 PLSMFRRHTVYLDSYAVINDLSTKNEGTRLAIKALELRFHGAKVVSCLAEGVSHVIIGEDHSRVADFKAFRRTF------
At   802 ---LLSCCVYLYPYSQTESTEEEALLGIMAKRLMLEVLMAGGKVSNNLAH-ASHLVVLANAEEPLDFTLVSKSFSEMEKR

Sc   913 ---PKIARVVAPEWVEHSINENCQVPEEDFPVNY-----------------------------------------------
Hs   816 ---KRKFKILKESWVTDSIDK-CELQEENQYLI------------------------------------------------
At   878 LLLKKPLHVVSSHWLEESLQR-EEKLCEDVYTIRPKYMEESDTEESDKSEHDTTEVASQGSAQTKEPASSKIAITSSRGR

Sc       --------------------------------------------------------------------------------
Hs       --------------------------------------------------------------------------------
At   957 SNTRAVKRGRSSTNSLQRVQRRRGKQPSKISGDETEESDASEEKVSTRLSDIAEETDSFGEAQRNSSRGKCAKRGKSRVG

Sc       --------------------------------------------------------------------------------
Hs       --------------------------------------------------------------------------------
At  1037 QTQRVQRSRRGKKAAKIGGDESDENDELDGNNNVSADAEEGNAAGRSVENEETREPDIAKYTESQQRDNTVAVEEALQDS

Sc       --------------------------------------------------------------------------------
Hs       --------------------------------------------------------------------------------
At  1117 RNAKTEMDMKEKLQIHEDPLQAMLMKMFPIPSQKTTETSNRTTGEYRKANVSGECESSEKRKLDAETDNTSVNAGAESDV

Sc       ------------------------
Hs       ------------------------
At  1197 VPPLVKKKKVSYRDVAGELLKDW
```

FIGURE 4



```
Sc    1 ------MSAIYKLSIQGIRSFDSNDRE--TIEFGKPLTLIVGMNGSGKTTIIECLKYATTGDLPPNS-KGGVFIHDPKIT
Hs    1 MLIFSVRDMFAKNSILGVRSFGIEDKDKQEITFFSPLTILVGPNGAGKTTIIECLKYICTGDFPPGT-KGNTFVHDPKVA
At    1 ------MSTVDKMLIKGIRSFDPENKN--VVTFFRPLTLIVGANGAGKTTIIECLKVSCTGELPPNARSGHSFIHDPKVA

Sc   72 GEKDIRAQVKLAFTSANGLNMIVTRNIQLLMKKTTTTFKTLEGQLVAINNS-GDRSTLSTRSLELDAQVPLYLGVPKAIL
Hs   80 QETDVRAQIRLQFRDVNGELIAVQRSMVCTQKSKKTEFKTLEGVITRT-KH-GEKVSLSSKCAEIDREMISSLGVSKAVL
At   73 GETETKACIKLRFKTAAGKDVVCIRSFQLTQKASKMEYKAIESVIQTINPHTGEKVCLSYRCADMDREIPALMGVSKAIL

Sc  151 EYVIFCHQEDSLWPLSEPSNLKKKFDEIFQAMKFTKALDNLKSIKKDMSVDIKLLKQSVEHLKLDKDRSKAMKLNIHQLQ
Hs  158 NNVIFCHQEDSNWPLSEGKALKQKFDEIFSATRYIKALETLRQVRQTQGQKVKEYQMELKYLKQYKEKACEIRDQITSKE
At  153 ENVIFVHQDESNWPLQDPSTLKKKFDDIFSATRYTKALEVIKKLHKDQAQEIKTFKLKLENLQTLKDAAYKLRESIAQDQ

Sc  231 TKIEQYNEEVSEIESQLNEITEKSDKLFKSNQDFQKILSKVENLKNTKLSIS-DQVKRLSNSIDILDLSKPDLQNLLANF
Hs  238 AQLTSSKEIVKSYENELDPLKNRLKEIEHNLSKIMKLDNEIKALDSRKKQMEKDNSELEEKMEKVFQGTDEQLNDLYHNH
At  233 ERTESSKVQMLELETSVQKVDAEVHNKEMMLKDLRKLQDQVSIKTAEPSTLFKEQQRQYAALPEENEDTIEELKEWKSKF

Sc  310 SKVEMDRNNQLRDLETDISSLKDRQSSLQSLSNSLIRRQGELEAGKETYEKN-RNHLSSLKEAFQHKFQGLSNIENSDMA
Hs  318 QRTVREKERKLVDCHRELEKLNKESRLLNQEKSELLVEQGRLQLQADRHQEHIRARDSLIQSLATQLELDGFERGPFSER
At  313 EERLALLGTKIRKMEREMVDTETTISSLHNAKTNYMLEISKLQTEAEAHMLLKNERDSTIQNIFFHYNLGNVPSTPFSTE

Sc  389 QVNHEMSQFKAFISQDLTDTIEQFAKDIQLKETNLSDLIKSITVDSQNLEY-NKKDRSKLIHDS--EELAEKLKSFKSLS
Hs  398 QIKNFHKLVRERQ-EGEAKTANQLMNDFAEKETLKQKQIDEIRDKKTGLGR-IIELKSEILSKKQNELKNVKYELQQLEG
At  393 VVLNLTNRIKSRVGELEMDLLDKKKSNETALSTAWDCYMDANDRWKSIEAQKRAKDEIKNGISKRIEEKEIERDSFEFEI

Sc  466 IQDSLNHELENLKTYKEKLQSWESENIIPKLNQKIEEKNNEMIILENQIEKFQDRIMKTNQQADLYAKLGLIKKSINTKL
Hs  476 SSDRILELDQELIKAERELSKAEKNSNVETLKMEVISLQNEKADLDRTLRKLDQEMEQLNHHTTTRTQMEMLTKDKADKD
At  473 STVDVKQTDERELQVQVELERKTKQNSERGFESKIEQKQHEIYSLEHKISTLNRERDVMAGDAEDR-LLTRIDECKDRIR

Sc  546 DELQKITEKLQNDSRIRQVFPLTQEFQRADLEMDFQKLFINMQKNIAINNKKIHELDRRYTNALYNLNTIEKDLQDNQKS
Hs  556 EQIRKIKSRHSD-----ELTSLLGYFPNKKQLEDWLHSKS---KEINQTRDRLAKLNKELASSEQNKNHINNELKRKEEQ
At  552 GVLKGRLPPEKD-----MKREIVQALRSIEREYDDLSLKS---REAEKEVNMLQMKIQEVNNS--LFKHNKDTESRKRYI

Sc  626 KEKVIQLLSENEPEDCTIDEYNDVLEETELSYKTALENLKMHQTTLEENRKALEIAERDSCCYLCSRKFE--NESFKSKL
Hs  628 LSSYEDKLFDVCGSQDFESDLDRLKEEIEKSSKQRAMLAGATAVYSQFITQLTDENQS--CCPVCQRVFQTEAELQEAIS
At  622 ESKLQALKQESVTIDAYPKLLESAKDKRDDRKREYNMANGMRQMFEPFEKRARQEHS----CPCCERSFT-ADEEASFIK

Sc  704 LQEEKTKIDENFEKLLKDTVQNEKEYLHSLRLLEKHIITLNSIN-EKIDNSQKCLEKAKEETKTSKSKLDELEVDSTKLK
Hs  706 DLQSKLRLAPDKLKSTESELKKKEKRRDEMLGLAPMRQSIIDLKEKEIPELRNKLQNVNRDIQRLKNDIEEQETLLGTIM
At  697 KQRVKASSTGEHLKALAVESSNAESVFQQLDKLRAVFEEYSKLTTEIIPLAEKTLQEHTEELGQKSEALDDVLGISAQIK

Sc  783 DEKELAESEIRPLIEKFTYLEKELKDLENSSKTISEELSIYNTSEDGIQTVDELRDQQRKMNDSLRELRKTISDLQMEKD
Hs  786 PEEESAKVCLTD-VTIMERFQMELKDVERKIAQQAAKLQG-IDLDRTVQQVNQEKQEKQHKLDTVSSKIELNRKLIQDQQ
At  777 ADKDSIEALVQP-IENADRLFQEIVSYQKQIEDLEYKLDFRGLGVKTMEEIQSELSSLQSSKDKLHGELEKLRDDQIYME

Sc  863 EKVRENSRMINLIKEKELTVSEIESSLTQKQNIDDSIRSKRENINDIDSRVKELEARIISLKNKKDEAQSVLDKVKNERD
Hs  864 EQIQHLKSTTNELKSEKLQISTNLQRRQQ---LEEQTVELSTEVQSLYREIKDAKEQVSPLETTLEKFQQEKEELINKKN
At  856 RDISCLQARWHAVREEKAKAANLLRDVTK---AEEDLERLAEEKSQLDLDVKYLTEALGPLSKEKEQLLSDYNDMKIERN

Sc  943 IQVENKQKTLADINRLIDRFQTIYNEVVDFEAKGFDELQTTIKELELNK-----AQMLELKEQLDLKSNEVNEERRKLAD
Hs  941 TSNKIAQDKLNDIKEKVKNIHGYMKDIENHIQDGKDDYMKQKETELNK----VIAQLSECEKHKEKINEDMRLMRQDIDT
At  933 QEYEELAEKKRNYQQEVEALLKASYKINDCFTRYHDLKKGERLDDIQEKQRLSDSQLQSCEARKNELAGELNRNKDLMRN

Sc 1018 SNIEEKNLKQNLELIELKSQLQHIESEISRLDVQNA-EAERDKYQEESLRLRTRFEKLSSENAGKLGEMKQLQNQIDSLT
Hs 1017 QKIQERWLQDNLTLRKRNEELKEVEEEGKQHLKEMG-QMQVLQMKSEHQKLEENIDNIKRNHNLALGRQKGYEEEIIHFK
At 1013 QDQLRRNIEDNLNYRTTKAKVEELTREIESLEEQILNIGGIAAVEAEIVKILRERERLLSELNRCRGTVSVYESSISKNR

Sc 1097 HQLR-TDYKDIEKNYHKEWVELQTRSFVTDDIDVYSKALDSAIMKYHGLKMQDINRIIDELWKRTYSGTDIDTIKIRSDE
Hs 1096 KELREPQFRDAEEKYREMMIVMRTTELVNKDLDIYYKTLDQAIMKFHSMKMEEINKIIRDLWRSTYRGQDIEYIEIRSDA
At 1093 VELKQAQYKDIDKRHFDQLIQLKTTEMANKDLDRYYNALDKALMRFHTMKMEEINKIIRELWQQTYRGQDMDYIRIHSDS

Sc 1176 VS---STVKGKSYNYRVVMYKQDVELDMRGRCSAGQKVLASIIIRLALSETFGANCGVIALDEPTTNLDEENIESLAKSL
Hs 1176 DENVSASDKRRNYNYRVVMLKGDTALDMRGRCSAGQKVLASLIIRLALAETFCLNCGIIALDEPTTNLDRENIESLAHAL
At 1173 EG------AGTRSYSYKVVMQTGDTELEMRGRCSAGQKVLASLIIRLALAETFCLNCGILALDEPTTNLDGPNSESLAGAL

Sc 1253 HNIINMRRHQKNFQLIVITHDEKFIGHVNAAAFTDHFFKVKRDDRQKSQIEWVDINRVTY---
Hs 1256 VEIIKSRSQQRNFQLLVITHDEDFVELLGRSEYVEKFYRIKKNIDQCSEIVKCSVSSLGFNVH
At 1248 LRIMEDRKGQENFQLIVITHDERFAQMIGQRQHAEKYYRVAKDDM------------------
```

BEST AVAILABLE COPY